

textos para discussão

139 | Junho de 2019

Conceitos e metodologias de avaliação aplicáveis ao financiamento de exportações de serviços

Rafael de Carvalho Cayres Pinto
Elydia Silva Hirata
Luiz Filipe de Castro Neves

Presidente do BNDES

Joaquim Levy

Diretoria de Empresas

Denise Pavarina

Área de Comércio Exterior

Leonardo Pereira Rodrigues dos Santos

textos para discussão

139 | Junho de 2019

**Conceitos e metodologias
de avaliação aplicáveis
ao financiamento de
exportações de serviços**

**Rafael de Carvalho Cayres Pinto
Elydia Silva Hirata
Luiz Filipe de Castro Neves**

Resumo

Este trabalho busca introduzir sucintamente os elementos pertinentes ao tema “avaliação de políticas públicas”, apresentando os principais conceitos e metodologias de estimação compatíveis com as hipóteses/situações a serem avaliadas. Adicionalmente, discute como essas metodologias podem ser aplicadas às avaliações de efetividade/impacto previstas para as operações de financiamento realizadas pelo BNDES no âmbito da linha Exim Pós-embarque Serviços. Trata-se da primeira parte do trabalho de avaliação dos impactos e resultados das operações de financiamento às exportações do Banco.

Palavras-chave: Monitoramento e avaliação. Efetividade. Avaliação de programas. Inferência causal. Estimação de efeitos de tratamento. Financiamento à exportação. Exportação de serviços.

Abstract

The goal of this paper is to describe the basic elements in the field of public policy evaluation, presenting the main concepts and estimation methods compatible with the hypotheses/situations to be assessed. In addition, the paper discusses how these methods can be applied in the prospective effectiveness/impact evaluations of the export credit support provided by BNDES under its Exim Post-shipment Services financing. This is the first part of an ongoing process of internal evaluation of the impacts and results of the Bank's export credit support.

Keywords: Monitoring and evaluation. Effectiveness. Program evaluation. Causal inference. Estimation of treatment effects. Export finance. Service exports.

Sumário

1. Introdução	9
2. Aspectos conceituais e metodológicos pertinentes ao tema “avaliação”	10
2.1 Aspectos conceituais	10
2.2 Métodos quantitativos	13
2.2.1 Métodos baseados em dados experimentais	14
2.2.2 Métodos baseados em dados observacionais	14
2.2.3 Métodos baseados em dados quase-experimentais	17
3. Monitoramento e avaliação da linha de financiamento Pós-embarque Serviços	19
3.1 Atividades de monitoramento e avaliação da linha Pós-embarque Serviços	21
3.1.1 Elaboração dos indicadores de eficácia e efetividade a serem avaliados e monitorados	21
3.1.2 Avaliação <i>ex post</i> da efetividade do apoio do BNDES diante dos objetivos da linha Pós-embarque Serviços	22
3.2 Metodologias aplicáveis à avaliação de impacto da linha Pós-embarque Serviços	23
3.3 Dados	25
Referências	26
Apêndice metodológico – identificação e estimação do efeito do tratamento	29
Descrição geral e notação	29
Parâmetro de interesse	30
Dados experimentais	31
Seleção em observáveis	32
Seleção em não observáveis	36

Dados dinâmicos: diferenças em diferenças e controle sintético	36
Diferenças em diferenças	36
Controle sintético	38
Contextos quase-experimentais: variáveis instrumentais e <i>regression discontinuity design</i>	40
Variáveis instrumentais	40
<i>Regression discontinuity design</i>	42
Referências	44

Rafael de Carvalho Cayres Pinto é economista na Área de Comércio Exterior do BNDES; Elydia Silva Hirata é coordenadora de serviços na Área de Comércio Exterior do BNDES; e Luiz Filipe de Castro Neves, gerente na Área de Comércio Exterior do BNDES. Os autores agradecem a Caroline Medeiros dos Santos e Isabella Miranda Meyer a assistência na elaboração deste trabalho, e a Andressa Biato Senra, Carlos Frederico Braz de Souza, Leonardo Pereira Rodrigues dos Santos, Marcelo Oliveira Santos e Priscilla Assis Pinto da Matta os comentários e sugestões. Erros e omissões remanescentes são de inteira responsabilidade dos autores.

1. Introdução

O processo de monitoramento e avaliação de impacto tem se difundido e desenvolvido nas últimas décadas. Sua importância reside na disponibilização de informações que subsidiem decisões a respeito de aprimoramento, reformulação ou expansão de programas avaliados. Para além das aplicações no setor privado, processos de monitoramento e avaliação têm ocupado seu espaço na escolha e no aprimoramento de políticas públicas, em um ambiente de maior busca por transparência e prestação de contas dos governos perante a sociedade. Embora sempre necessária, essa discussão ganha fôlego todas as vezes em que se aponta para a necessidade de racionalização dos gastos públicos, com vistas a aumentar a eficiência e eficácia das ações do setor público.

Nos últimos anos, o BNDES tem buscado implantar procedimentos de avaliação sistemática de seus programas de financiamentos como forma de aprimorar a aplicação dos recursos públicos que lhe são confiados, bem como para responder à crescente demanda da sociedade e dos órgãos de controle por informações a respeito dos benefícios sociais gerados na aplicação desses recursos. Nesse sentido, há, institucionalmente, interesse cada vez maior sobre o tema, com o desenvolvimento de metodologias para avaliação *ex ante* e *ex post* dos financiamentos solicitados e concedidos. Os financiamentos à exportação integram esse processo, e, para a Área de Comércio Exterior, estão previstas duas avaliações de efetividade referentes à linha BNDES Exim Pós-embarque Serviços, responsável pelo financiamento às exportações brasileiras destinadas a obras de infraestrutura no exterior.

Dessa forma, mostra-se interessante apresentar os principais elementos pertinentes ao tema “avaliação de políticas públicas”, introduzindo sucintamente os principais conceitos e metodologias de estimação de impacto compatíveis com as hipóteses/situações a serem avaliadas. Adicionalmente, este trabalho exemplifica formas de aplicação dessas metodologias às avaliações de efetividade previstas para as operações já financiadas pela linha Exim Pós-embarque Serviços. Tais conceitos e metodologias, no entanto, são aplicáveis às diversas linhas e programas do BNDES.

Além desta introdução, o presente texto se encontra dividido em outras duas seções, além de um apêndice a respeito da metodologia utilizada. A próxima seção introduz brevemente os aspectos conceituais e os principais métodos estatísticos e econométricos utilizados em avaliações de impactos. A terceira seção apresenta um pequeno histórico da linha BNDES Exim Pós-embarque Serviços e discute a aplicação das metodologias apresentadas para os financiamentos dessa linha. O apêndice metodológico retoma a introdução dos métodos quantitativos da segunda seção com maior nível de formalidade, utilizando conceitos de probabilidade e estatística.

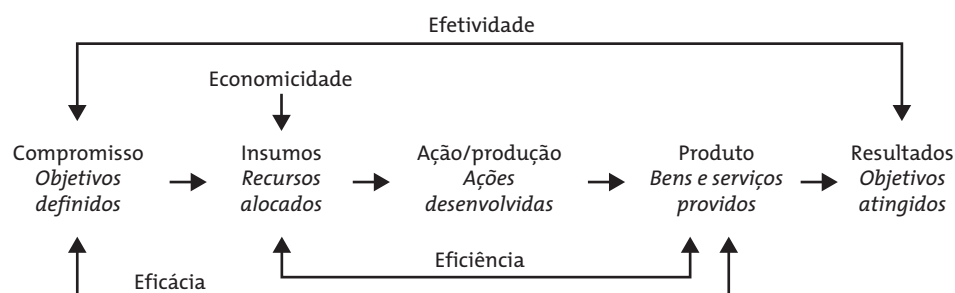
2. Aspectos conceituais e metodológicos pertinentes ao tema “avaliação”

2.1 Aspectos conceituais

Avaliação é uma aferição objetiva da extensão do cumprimento dos objetivos de uma intervenção, política pública ou atividade. Tem como finalidades contribuir para o aprimoramento do planejamento e da gestão e demonstrar os resultados alcançados para as partes interessadas.

Entre as dimensões de desempenho passíveis de avaliação, distinguem-se: economicidade, eficiência, eficácia e efetividade. O significado de cada uma dessas dimensões corresponde a diferentes etapas do modelo lógico¹ da intervenção (Figura 1), desde o **compromisso** (definição do objetivo) até os **resultados** obtidos, incluindo a aquisição de **insumos** a serem empregados na ação ou produção e que se materializa na entrega dos *produtos*, cuja provisão deve contribuir para o alcance dos **resultados**.

Figura 1. Modelo lógico de uma atividade



Fonte: TCU (2010), adaptado do padrão internacional ISSAI 3000 (INTOSAI, 2004).

A economicidade e a eficiência dizem respeito à minimização de custos, respectivamente, na **aquisição** e na **utilização** dos insumos. Eficácia e efetividade, por sua vez, correspondem à capacidade da intervenção de gerar, respectivamente, os produtos (objetivos intermediários) e resultados (objetivos finais) pretendidos, independentemente do custo incorrido. Assim, verifica-se a:

- **Economicidade**, quando a organização adquire os insumos ao menor preço possível e na quantidade estritamente necessária para a realização da atividade – respeitando o padrão de qualidade almejado.
- **Eficiência**, quando o processo maximiza a relação entre a quantidade de produtos obtida e os insumos utilizados; ou seja, obtém uma relação ótima entre insumos e produtos.

¹ Também conhecido na literatura de avaliação de programas como modelo insumo-produto. Não confundir com o modelo econômico de insumo-produto de Leontief, que é utilizado – inclusive no BNDES [ver, por exemplo, Najberg e Ikeda (1999)] – para avaliar efeitos de encadeamentos econômicos sobre produção, emprego e renda.

- **Eficácia**, quando se verifica a capacidade da intervenção em entregar os produtos julgados necessários para a obtenção dos objetivos.
- **Efetividade**, quando a intervenção gera a mudança social pretendida, ou seja, é capaz de materializar os resultados almejados ou objetivos finais.

Os conceitos discutidos anteriormente podem ser associados a duas perguntas fundamentais (INTOSAI, 2004): (i) as coisas estão sendo feitas da forma correta? e (ii) as coisas certas estão sendo feitas? A primeira diz respeito ao desempenho de uma organização vista como um “produtor” no contexto de uma atividade, gerenciando adequadamente os recursos à sua disposição e tomando as melhores decisões possíveis. Refere-se, basicamente, à análise da economicidade e da eficiência. A segunda questiona o mérito da realização da atividade. Trata-se de avaliar se a atividade foi concebida de forma adequada para o alcance de sua finalidade; ou seja, corresponde às questões da eficácia e da efetividade.

A segunda questão, em certo sentido, tem um caráter mais essencial que o da primeira. Um programa efetivo que não tenha eficiência ou economicidade pode ser reformado de modo a se reduzir desperdícios. No entanto, a falta de efetividade evidencia que algum elo da lógica de intervenção não funciona como previsto.²

Outros conceitos aplicáveis à avaliação de uma política pública são:

- **Impacto**: quando usado em contraste com o conceito de resultado, refere-se particularmente a efeitos de longo prazo associados aos objetivos mais amplos da política pública. Neste caso, resultado passa a significar o alcance de objetivos de curto e médio prazo especificamente almejados pela intervenção, e impacto refere-se ao alcance de objetivos gerais de longo prazo (para os quais a intervenção específica contribui, mas não é necessariamente capaz de atender).
- **Efeito direto versus indireto**: equivale à dicotomia resultado *versus* impacto.
- **Efeito causal**: efeito líquido de um programa ou atividade, excluída a contribuição de fatores externos. Implica a comparação da mudança observada ao longo do período da intervenção com uma estimativa do que aconteceria caso o programa não existisse (o “resultado contrafactual”). Trata-se de um requisito necessário a qualquer avaliação de efetividade.
- **Externalidade**: termo utilizado pela economia para designar efeitos de uma decisão sobre indivíduos que não participam da tomada dessa decisão.

² Neste caso, a explicitação da cadeia de efeitos e premissas em um quadro lógico pode auxiliar na detecção de eventuais falhas na concepção da intervenção e, portanto, indicar caminhos para sua correção.

No contexto de avaliação, aplica-se aos efeitos colaterais sobre agentes sociais que não são beneficiários imediatos da intervenção.

- **Sustentabilidade:** capacidade de um programa gerar efeitos de longo prazo; ou seja, as mudanças alcançadas tendem a se perpetuar no tempo (dinâmica instituída permanece depois do programa).

Quanto ao momento em que ocorrem, em relação à intervenção ou ao programa, as avaliações podem ser classificadas como:

- **Ex ante:** ocorre anteriormente à intervenção e busca, com base em uma série de atributos e expectativas, mapear os efeitos esperados dessa intervenção e, assim, selecionar as melhores estratégias ou unidades para sua execução. Embora apoiadas tanto quanto possível em evidências empíricas, avaliações deste tipo recorrem a modelos teóricos para transpor os dados históricos para a realidade da intervenção analisada.
- **Ex post:** objetiva a verificação da efetiva realização dos resultados planejados/almejados. São de caráter predominantemente empírico, tendo como base a inferência estatística sobre os resultados observados, de modo a verificar em que medida as metas pretendidas foram alcançadas.

Idealmente, as avaliações *ex ante* e *ex post* devem estar integradas, com a primeira fornecendo os elementos e as expectativas a serem confirmados ou refutados pela segunda. Dessa forma, pode-se determinar o que se espera do programa e verificar se o que havia sido previsto de fato ocorreu. Além disso, durante a intervenção, é importante monitorar e analisar os indicadores selecionados, para uma eventual correção de rumo do projeto, antes de seu encerramento.

Outros dois conceitos interessantes para a geração e aferição de resultados de uma política pública referem-se à amplitude/ao alcance dessa política. São eles:

- **Focalização:** refere-se à alocação dos recursos disponíveis de um orçamento nas unidades ou atividades que possibilitarão melhores resultados e/ou em que eles são mais necessários. Por exemplo, nos grupos mais vulneráveis ou nas falhas de mercados. A focalização tende a aumentar a eficiência da política pública; contudo, um programa muito focalizado pode ter um efeito total restrito.³ Uma das dificuldades da focalização é o custo ou a disponibilidade de obter informações que permitam identificar as melhores unidades a receber a intervenção.
- **Cobertura:** refere-se ao percentual de unidades que participam da intervenção em relação ao universo de unidades que atendem aos critérios de

³ Por exemplo, no caso das exportações, as políticas do BNDES são focalizadas nos setores de alto valor agregado.

participação. Supondo que a intervenção esteja corretamente focalizada, espera-se que o aumento da cobertura eleve o impacto total, mas reduza seu impacto marginal, uma vez que cada nova unidade atendida será beneficiada em menor grau. Acima de um determinado patamar, a expansão do programa poderá até mesmo tornar-se inócua.

2.2 Métodos quantitativos

Avaliar o impacto de uma intervenção ou **tratamento** sobre os resultados almejados consiste em comparar o desempenho das empresas, indivíduos, mercados, sociedades etc. – genericamente **unidades** – sob efeito dessa intervenção com o que ocorreria em sua ausência. Tecnicamente, trata-se de uma inferência estatística a respeito das distribuições de probabilidade dos valores de indicadores considerados relevantes – as **variáveis de interesse** – sob as duas condições possíveis: presença e ausência do tratamento.

A dificuldade fundamental dessa inferência decorre da impossibilidade de observar o desempenho de uma mesma unidade sob ambas as condições, uma vez que uma dessas situações é **contrafactual**, isto é, representa uma possibilidade não verificada na realidade. Em outras palavras, só se observam os resultados alcançados sob o tratamento para as unidades tratadas e o resultado obtido na ausência do tratamento para as que não foram tratadas.⁴

Assim, para estimar os resultados (contrafactuais) que seriam obtidos pelos beneficiários da intervenção – chamados de unidades **tratadas** ou **grupo de tratamento** –, caso não estivessem sujeitas a ela, deve-se recorrer a observações de unidades que de fato não foram alvo da intervenção. Esse segundo conjunto de unidades, denominado **grupo de controle**, deve ser suficientemente semelhante ao grupo de tratamento, de modo que seu desempenho forneça informação válida sobre o desempenho contrafactual do grupo de tratamento.

A literatura de estimação de efeitos de tratamento conta com uma série de métodos que buscam essencialmente: (i) identificar rigorosamente o **efeito** da intervenção, ou seja, sua relação **causal** com a variável de interesse, isolando-o de outros fatores; e (ii) evitar a ocorrência de **viés** sistemático, isto é, tendência estatística à sobre-estimação ou subestimação. Essa variedade de métodos visa atender aos requisitos mencionados sob diversas hipóteses referentes à natureza dos dados, especialmente quanto à relação entre os resultados potenciais (real e contrafactual) e a atribuição aos grupos de controle e tratamento.

⁴ Segundo Khandker, Koolwal e Samad (2010, p. 25), “an impact evaluation is essentially a problem of missing data, because one cannot observe the outcomes of program participants had they not been beneficiaries”.

2.2.1 *Métodos baseados em dados experimentais*

Do ponto de vista da estimação do impacto, a situação mais simples ocorre quando a incidência da intervenção segue um **desenho experimental**, ou seja, a seleção das unidades beneficiadas pela intervenção ocorre de forma aleatória dentro de um conjunto inicial de beneficiários potenciais.⁵ Nesse caso, o conjunto das unidades não selecionadas constitui naturalmente um grupo de controle apropriado. De fato, a simples comparação entre os desempenhos dos grupos quanto às variáveis relevantes é adequada, uma vez que não há diferença estatística significativa quanto a outros fatores que pudessem afetar esses desempenhos.

O estabelecimento de um desenho experimental permite obter estimativas robustas para o impacto do programa de forma computacionalmente simples e praticamente sem necessidade de impor hipóteses. No entanto, esse método nem sempre é factível, em virtude de questões éticas (implica tratar diferentemente unidades idênticas), de custos (obter e gerar informações suficientes para assegurar que a atribuição do tratamento foi de fato aleatória) e de viabilidade política (deixar de atender a parte do público-alvo quando se acredita que a intervenção é benéfica).

Além disso, mesmo sob seleção aleatória, é preciso atentar para fenômenos que podem comprometer a inferência. Destacam-se os seguintes problemas:

- **Validade externa**, que significa que a inferência obtida no âmbito do experimento não se aplica para a população em geral. Isso ocorre quando o grupo de participantes do experimento não é representativo da população (isto é, tem características estatisticamente distintas).
- **Externalidades** entre as unidades, situação em que o programa também afeta (indiretamente) o grupo de controle, viesando a comparação.
- **Contaminação de amostra**, quando a incidência do tratamento não segue exatamente o previsto pela seleção aleatória. A contaminação pode ocorrer quando a participação de fato depende da vontade das unidades. Assim, uma unidade pode rejeitar o tratamento mesmo quando selecionada para participar ou se beneficiar do programa mesmo quando não selecionada.⁶

2.2.2 *Métodos baseados em dados observacionais*

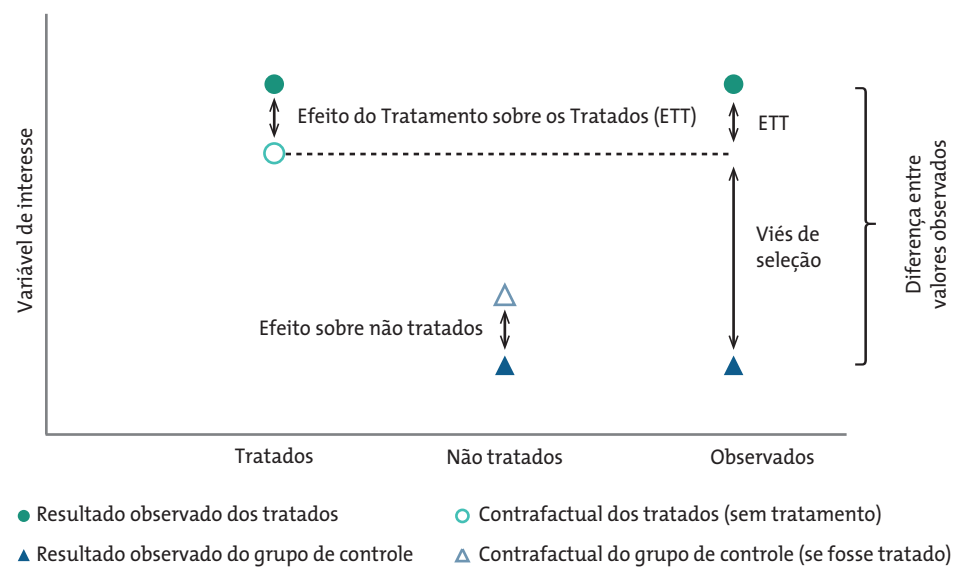
Quando não se dispõe de dados experimentais, há o risco de que a diferença de desempenho observada tenha origem em outros fatores além da intervenção em tela. Nesse caso, a comparação simples é sujeita a diversas fontes de viés sistemático,

⁵ Geralmente participantes de uma versão-piloto do programa em tela.

⁶ Por exemplo, se a intervenção consiste na provisão pública de um curso de capacitação, um participante selecionado pode faltar ao curso e um não selecionado pode – tendo, em função do programa, tomado ciência da existência da intervenção – buscar uma formação similar de forma privada.

em particular o conhecido como **viés de seleção**, ilustrado na Figura 2. Este decorre do fato de que a seleção para o tratamento geralmente depende de escolhas dos operadores da própria intervenção e dos potenciais participantes, resultando em relações estatísticas entre participação, resultados potenciais e outras variáveis que estão, *a priori*, fora do controle da equipe de avaliação. O ajuste necessário para que uma comparação de desempenhos revele o verdadeiro efeito da intervenção depende, então, de hipóteses sobre essas relações.

Figura 2. Dados observacionais



Fonte: Elaboração própria.

Um contexto relativamente brando é descrito pela **hipótese de seleção sobre observáveis**, que postula que as diferenças relevantes entre as unidades estão expressas em características observadas pela equipe de avaliação, denominadas genericamente como **variáveis auxiliares ou de controle**. Nesse caso, é lícito comparar unidades com o mesmo nível das variáveis auxiliares, e está disponível a seguinte variedade de técnicas que permitem obter inferências adequadas:

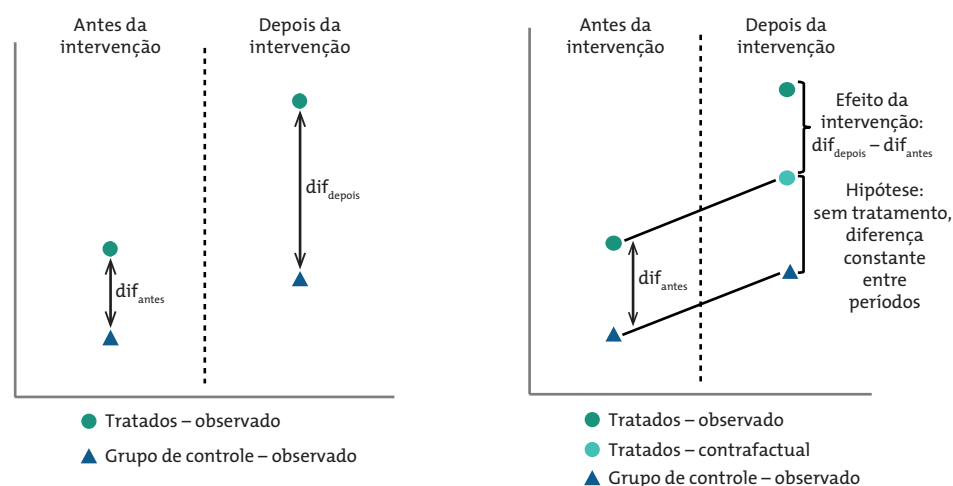
- **Regressão com variáveis de controle:** a estimação do resultado consiste em comparar o resultado observado das unidades tratadas com um contrafactual obtido por meio da assimilação da relação estatística entre as características observáveis e a variável de desempenho no grupo de controle.
- **Matching:** pareamento de unidades tratadas com membros similares do grupo de controle, seguido de comparação de desempenho das primeiras com as segundas. Em outras palavras, apenas o desempenho das unidades de controle escolhidas como “pares” de unidades de tratamento é utilizado na comparação.

- **Balanceamento por *propensity score*:** quando há uma relação estatística entre as características observáveis e a probabilidade de cada unidade pertencer a um dos dois grupos, é possível utilizar essa relação para reponderar/rebalancear as observações do grupo de controle disponível, tornando-o comparável ao de tratamento.

Um obstáculo mais difícil de contornar é a possibilidade de que as unidades beneficiadas e não beneficiadas sejam estatisticamente distintas em aspectos não observáveis (diz-se que há **seleção em características não observáveis**) que afetam o desempenho. Essa possibilidade é bastante plausível, uma vez que os agentes econômicos geralmente dispõem de informações privadas – portanto indisponíveis para os observadores externos – e que podem afetar seu desempenho em diferentes circunstâncias. Quando isso ocorre, a observação do desempenho prévio ao efeito da intervenção pode ajudar a eliminar o viés. Nesse sentido, duas técnicas aplicáveis são:

- **Diferenças em diferenças (*diff in diff* ou método de efeitos fixos):** se a diferença dos fatores não observáveis entre os grupos permanece estável entre um período anterior e outro posterior à intervenção, é possível comparar a evolução de desempenho dos grupos entre esses períodos. Como por hipótese a diferença devida aos demais fatores não varia no tempo, pode-se atribuir a diferença nas evoluções dos dois grupos à intervenção realizada. Esse princípio é ilustrado na Figura 3.

Figura 3. Diferenças em diferenças

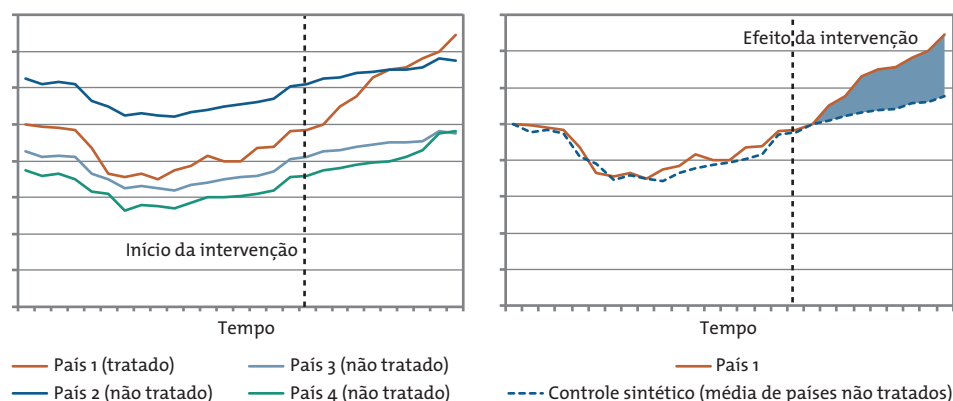


Fonte: Elaboração própria.

- **Controle sintético:** é uma generalização do *diff in diff* que permite que o efeito dos fatores não observáveis entre os grupos varie ao longo do tempo. O método é indicado para análises de tratamentos que afetem um pequeno número de unidades, como países ou setores econômicos, para as quais se disponha de

informações de múltiplos períodos de tempo antes e depois da intervenção. Ele permite controlar o efeito das diferenças tanto nas características observáveis quanto nas não observáveis, na medida em que estas últimas refletem na trajetória das unidades analisadas no período próximo ao da intervenção. As informações de vários períodos de tempo permitem determinar o melhor grupo de comparação para cada dimensão a ser analisada. Em resumo, a metodologia baseia-se na construção de uma unidade hipotética composta por uma média ponderada das principais características e resultados de unidades de controle, previamente à intervenção. Na prática, trata-se de selecionar um número de unidades não beneficiadas e atribuir pesos a elas, a fim de gerar o desempenho de uma unidade hipotética com características observáveis, bem como uma trajetória das variáveis de desempenho pré-intervenção similares às de uma unidade beneficiada, como representado na Figura 4. Uma das vantagens do controle sintético é retirar a subjetividade/arbitrariedade da escolha das unidades de comparação, que passam a ser orientadas pelos dados históricos das potenciais unidades de controle disponíveis, como apontado por Abadie e Gardeazabal (2003) e Abadie, Diamond e Hainmueller (2010).

Figura 4. Controle sintético



Fonte: Elaboração própria.

2.2.3 Métodos baseados em dados quase-experimentais

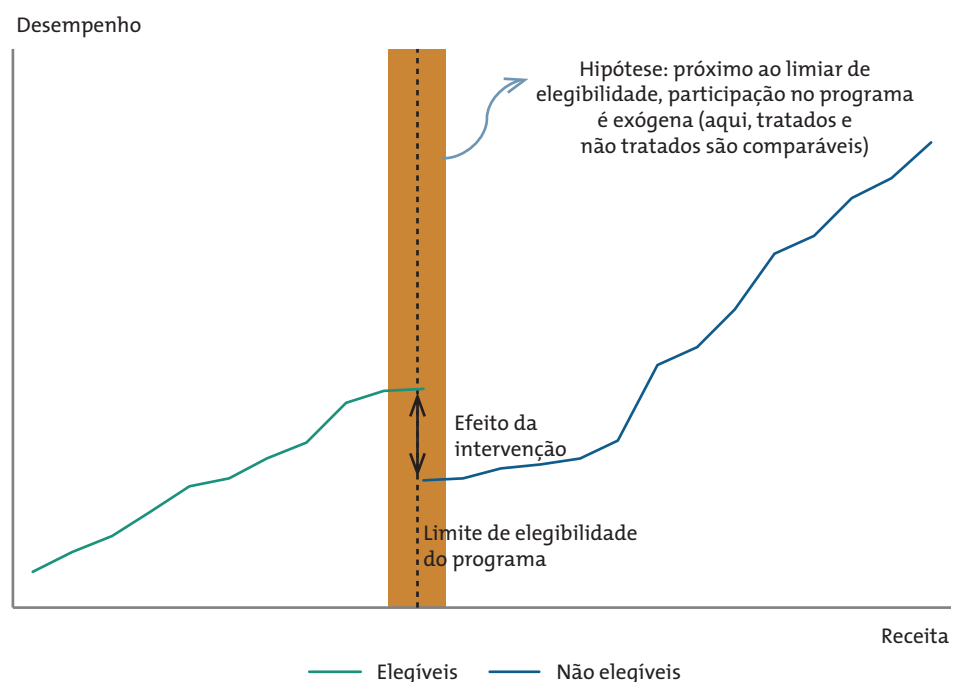
Outra forma de contornar o problema de viés de seleção é considerar o efeito de variações que criam circunstâncias equivalentes às de um experimento. Exemplos de metodologias são:

- **Variável instrumental:** para uma determinada intervenção é uma característica correlacionada com a probabilidade de uma unidade passar por intervenção, mas não correlacionada com fatores não observáveis que afetam seu desempenho. Nesse caso, é possível realizar inferência apropriada com base na comparação do desempenho entre empresas com diferentes níveis da variável instrumental (com o devido controle dos demais fatores observados). Uma possível fonte de variáveis instrumentais é a ocorrência de um experimento

natural, que consiste na existência de um fator aleatório exógeno (isto é, fora do controle das empresas) que afeta a probabilidade de incidência da intervenção. Um exemplo de experimento natural seria o estabelecimento de condições de financiamento diferenciadas associadas a um recorte geográfico, que afetaria diferentemente empresas com diferente exposição à região afetada.

- **Regressão descontínua:** essa técnica explora uma variação descontínua da incidência da intervenção com respeito a um limiar conhecido de determinada *variável de atribuição* (conhecida na literatura como “*forcing*”, “*running*”, ou “*assignment-variable*”). Essa descontinuidade pode ser determinística (*sharp*), quando a descontinuidade determina a exposição ou não da unidade à intervenção, ou probabilística (*fuzzy*), quando afeta apenas a probabilidade de incidência. Desde que não ocorra “manipulação” da variável de atribuição – unidades ajustando o valor dessa variável para participar ou deixar de participar do programa (o que caracterizaria autoseleção) –, é possível considerar as unidades de controle ou tratamento próximas ao limiar como grupos comparáveis. Consequentemente, a diferença entre seus desempenhos consiste em um estimador não viesado do efeito do tratamento nesse conjunto, como mostrado na Figura 5. Cabe ressaltar que se trata, nesse caso, de um efeito **local** do tratamento, isto é, o efeito sobre um subconjunto da população originalmente estudada, a saber, as unidades com variável de atribuição próxima ao limiar. Em razão dessa qualificação, diz-se que o método carece de validade externa, pois não é lícito extrapolar a conclusão sobre o efeito do tratamento para unidades distantes do limiar.

Figura 5. Regressão descontínua



Fonte: Elaboração própria.

3. Monitoramento e avaliação da linha de financiamento Pós-embarque Serviços

Desde meados do século passado, o Brasil tem adotado mecanismos de promoção de exportações. Até a década de 1980, com a preponderância das políticas de substituição de importações, a necessidade de garantir importações prioritárias e de reforçar o balanço de pagamentos diante de crises impôs a instituição de mecanismos de promoção das exportações.⁷ Dentre os incentivos fiscais e creditícios criados até então, podem ser destacados: (i) a instituição do Fundo de Financiamento às Exportações (Finex) (1966), com as modalidades pré e pós-embarque, financiamento para capital de giro (1967) e a equalização da taxa de juros (1979); (ii) a criação do crédito-prêmio de Imposto sobre Produtos Industrializados (IPI) e Imposto sobre Circulação de Mercadorias ICM (1969-1970); e (iii) o estabelecimento do programa Befiex-Comissão (1972), para programas especiais de exportação.⁸

A partir da década de 1990, notadamente marcada pela abertura comercial e pela privatização de empresas públicas, a inserção do país no mercado internacional de forma mais competitiva passou a integrar os objetivos da política de comércio exterior. Um conjunto de iniciativas reformulou o sistema público de financiamento e seguro às exportações, com a finalidade de suprir a demanda dos exportadores por mecanismos de apoio financeiro de médio e longo prazo (PINTO *et al.*, 2008).

Datam dessa época: (i) a regulamentação do Programa de Financiamento às Exportações (Proex), com a retomada da política de equalização das taxas de juros; (ii) a implantação do Sistema Integrado de Comércio Exterior (Siscomex);⁹ e (iii) o início do financiamento às exportações pelo BNDES para operações de pré-embarque (1990) e pós-embarque (1991).¹⁰ A partir de 1997, as operações de financiamento às exportações de aeronaves e de serviços, em especial de serviços de engenharia e construção, passaram a integrar o escopo do BNDES Exim.

⁷ Segundo Pinheiro *et al.* (1993), o início dos incentivos às exportações remontaria à década de 1950, com o uso de expedientes cambiais para incentivar as exportações não tradicionais. No entanto, segundo os mesmos autores, o período de 1964 a 1974 foi um marco na atividade exportadora em razão das mudanças introduzidas pela política de incentivo às exportações.

⁸ Voltado para grandes empresas com projetos de longo prazo de exportação de manufaturados e vinculados à expansão da capacidade produtiva do país.

⁹ Trata-se de um sistema informatizado com objetivo de integrar as atividades de registro, acompanhamento e controle das operações de comércio exterior, mediante um fluxo único e automatizado de informações. Conforme informações do Ministério do Desenvolvimento, Indústria e Comércio Exterior, por meio do Siscomex é possível acompanhar tempestivamente a saída e o ingresso de mercadorias no país, uma vez que os órgãos de governo intervenientes no comércio exterior podem, em diversos níveis de acesso, controlar e interferir no processamento de operações para melhor gestão de processos. Por intermédio do próprio sistema, o exportador (ou o importador) trocam informações com os órgãos responsáveis pela autorização e fiscalização. Para mais informações sobre o tema, ver: <http://www.mdic.gov.br/Siscomex/Siscomex.html>. Acesso em: 24 abr. 2016.

¹⁰ Finamex, posteriormente substituído pelo BNDES Exim.

Como todas as ações do BNDES, o objetivo do apoio à exportação é a geração de emprego e renda no Brasil, o que ocorre por meio da ampliação e diversificação da base exportadora e do aumento da produção advinda da expansão dos mercados consumidores de produtos brasileiros. Além disso, as exportações trazem consigo o benefício de fortalecimento do balanço de pagamentos, ao garantir a entrada de divisas necessárias à manutenção do equilíbrio cambial e da capacidade de importação para a economia.

Nesse sentido, a linha BNDES Pós-embarque Serviços contribui para o objetivo do BNDES Exim, ao garantir um fluxo de longo prazo de entrada de divisas, possibilitar a participação brasileira em um setor competitivo (que movimenta mais de US\$ 500 bilhões no mundo)¹¹ e fomentar a cadeia de subfornecedores de bens e serviços, gerando uma rede de exportadores indiretos para as obras apoiadas.

Até o fim de 2016, a carteira de financiamento de serviços incluía 146 operações contratadas, em 15 países, com desembolsos de recursos no valor de US\$ 10 bilhões, dos quais cerca de 52% já haviam sido amortizados. Embora haja poucas empresas brasileiras concorrendo no mercado internacional de engenharia – caracterizado pela concentração em um pequeno número de grandes empresas e pela existência de barreiras à entrada –, foram identificados mais de quatro mil subfornecedores de bens e serviços nos projetos apoiados, dos quais dois terços se constituíam de pequenas e médias empresas.

Durante os quase vinte anos de existência, o financiamento do BNDES às exportações de serviços se ampliou e diversificou. O aumento da complexidade das operações apoiadas e o aprendizado institucional adquirido durante a realização dessas operações levaram à elaboração e ao aperfeiçoamento dos critérios adotados pela linha Pós-embarque Serviços, tendo seu último regulamento sido publicado em 2015. Como parte desse processo de aperfeiçoamento e atendimento à maior demanda de informação da sociedade e dos órgãos de controle do setor público, identificou-se a necessidade de implantação de um processo sistemático de monitoramento e avaliação de efetividade dessa linha, em consonância com as atividades previstas para outros financiamentos do BNDES. Dessa forma, uma série de atividades vem sendo desenvolvida pela Área de Comércio Exterior do BNDES, cujos objetivos se encontram listados a seguir.

¹¹ Receita internacional das 250 maiores empresas de construção do mundo, segundo publicações anuais da revista *Engineering News-Record* (ENR, 2017).

3.1 Atividades de monitoramento e avaliação da linha Pós-embarque Serviços

3.1.1 *Elaboração dos indicadores de eficácia e efetividade a serem avaliados e monitorados*

Um componente básico de qualquer sistema de monitoramento e avaliação é a escolha de indicadores que consistem em medições dos diversos elementos da lógica da intervenção (Figura 1). Em particular, os indicadores de eficácia e efetividade dizem respeito à medida dos produtos entregues e dos resultados atingidos. A formulação desses indicadores para a linha Pós-Embarque Serviços está relacionada a outras atividades no âmbito do BNDES.

A primeira se refere à Tese de Impacto do Investimento em Projetos (Tiip), metodologia de avaliação *ex ante* do impacto dos projetos a serem enquadrados,¹² elaborada pela Área de Planejamento em colaboração com as áreas operacionais e aprovada em 2017 pela Diretoria do BNDES.

A segunda se refere à elaboração do quadro lógico do apoio ao setor de engenharia e construção. Instrumento requerido pelo Tribunal de Contas da União (TCU), o quadro lógico delinea a lógica do apoio e, portanto, oferece uma estrutura para a escolha dos indicadores. Em um processo de avaliação de impacto, é muito importante que objetivos, metas e resultados esperados estejam bem definidos e integrados. A construção de um quadro lógico para um programa – ou de outras metodologias de organização lógica – impõe um esforço de articulação entre as diversas partes do processo, e seus resultados tendem a melhorar o desenho do programa e de sua posterior avaliação.

Uma ferramenta similar ao quadro lógico é o quadro de resultados, que explicita a cadeia de consequências almejadas em relação a produtos, resultados e impactos. Esse instrumento vem integrando os relatórios de análise de algumas áreas do BNDES desde janeiro de 2017. Em comparação ao quadro lógico, o quadro de resultados apresenta maior ênfase nos objetivos que se pretende atingir e envolve a quantificação dos resultados esperados.

Ou seja, os indicadores da linha Pós-embarque Serviços, a serem definidos e acompanhados, deverão: (i) estar em linha com os definidos pela Tiip, de modo a permitir a comparação do impacto pretendido com o obtido; (ii) compor o quadro lógico do apoio às exportações do setor de engenharia e construção; e (iii) subsidiar os quadros de resultados que integrarão os relatórios de análise de cada operação de financiamento.

¹² Etapa do processo de concessão de financiamento pelo BNDES, na qual se realiza uma análise técnica preliminar, em que se verifica a adequação do projeto às políticas operacionais e de crédito do Banco. Depois dessa análise, a operação é submetida à apreciação do Comitê de Enquadramento e Crédito (CEC) e, se aprovada, é “enquadrada” em alguma das linhas do BNDES.

3.1.2 *Avaliação ex post da efetividade do apoio do BNDES diante dos objetivos da linha Pós-embarque Serviços*

Para avaliar o desempenho histórico do financiamento às exportações de serviços do BNDES, estimando os benefícios sociais – sobretudo quanto à geração de emprego e renda – e eventuais externalidades, serão realizados dois estudos.

O primeiro terá como foco os objetivos de garantir a competitividade da empresa brasileira no exterior e de fortalecer o balanço de pagamentos. Esse estudo terá como objeto de análise os efeitos diretos do financiamento sobre o desempenho das empresas beneficiárias do apoio e sobre o balanço de pagamentos brasileiro, contemplando, então, os níveis de empresa e país. Nesse caso, a variável de tratamento estará associada, respectivamente, às contratações/liberações destinadas a cada empresa (nível empresa) e ao volume total do apoio (país).

O segundo estudo avaliará se foi alcançado o objetivo de fortalecimento da cadeia de fornecedores brasileiros e terá como objeto os efeitos indiretos dos financiamentos sobre as empresas cujos bens e serviços foram incorporados aos projetos apoiados (os **subfornecedores**). Assim, o nível de análise será o das empresas, e as unidades tratadas, em relação a algum ponto no tempo, serão aquelas que passaram a figurar como subfornecedores naquele momento.

De forma geral, esses estudos buscam: (i) compreender até que ponto a linha está alcançando seus objetivos e gerando os benefícios desejados para a economia brasileira; (ii) sintetizar as lições aprendidas, para utilização em novos projetos semelhantes; (iii) averiguar se a intervenção deveria ser acompanhada de outras iniciativas para geração de maior impacto; e (iv) indicar pontos de aprimoramento. De maneira específica, as avaliações buscarão responder às seguintes questões:

- **Setor de engenharia e construção e balanço de pagamentos:**
 - Houve fortalecimento da inserção e atuação das firmas brasileiras no exterior? As condições oferecidas foram compatíveis com as do mercado internacional? Em que medida?
 - Em que medida as exportações financiadas tiveram efeitos sobre o balanço de pagamentos brasileiro?
 - Houve impacto direto sobre o emprego no setor de engenharia e construção (geração e manutenção)? Em que medida?
- **Cadeia de subfornecedores:**
 - A cadeia de fornecedores foi fortalecida? O fornecimento de bens e serviços por subfornecedor brasileiro foi maior do que seria caso não houvesse financiamento?

- Houve geração de empregos diretos e indiretos? Em que medida?
- Quais lições foram aprendidas para utilização em novos projetos semelhantes? A intervenção deveria ser acompanhada de outras iniciativas para geração de maior impacto? O que poderia ser aprimorado na atuação do Banco?

3.2 Metodologias aplicáveis à avaliação de impacto da linha Pós-embarque Serviços

Na definição da metodologia de estimação do efeito do programa, objetiva-se encontrar o modelo que forneça os resultados mais próximos à realidade. Considerando a indisponibilidade dos resultados contrafactuais, essa escolha também envolve encontrar um grupo de controle apropriado.

Para tal, é importante não só conhecer o desenho da política pública (incluindo eventual viés de seleção para o tratamento), como também considerar a forma como os diferentes agentes respondem a ela (possibilidade de autosseleção) e a disponibilidade e seleção dos dados (viés de estimação derivado de cortes temporais, eliminação de unidades com informações faltantes etc.). O tratamento inadequado desses problemas pode enfraquecer e até mesmo invalidar a análise. Por exemplo, do ponto de vista do desenho da intervenção, considerando a natureza não aleatória da incidência do apoio da linha Pós-embarque Serviços e da escolha dos subfornecedores pelas empresas exportadoras (ou seja, a natureza não experimental dos dados), a comparação direta com o desempenho de empresas não beneficiadas não representa um procedimento adequado para a estimação do impacto desse apoio.

Por isso, antes de efetuar estimações, é indicada uma análise exploratória dos dados que permitam entender a dinâmica e os possíveis vieses do tratamento. Além das análises estatísticas tradicionais (média, variância, análise gráfica etc. aplicadas a características como tamanho, região, setor da firma etc.), serão acompanhadas as dinâmicas das variáveis de interesse nas firmas tratadas relativas a eventos tais como: passar a integrar a cadeia de subfornecedor de uma grande empresa, exportar pela primeira vez e obter/iniciar um projeto em determinado país.

Como já comentado, o planejamento do monitoramento e da avaliação de uma política pública deve, idealmente, ser realizado durante a concepção dessa mesma política pública. A prévia associação dos objetivos que se deseja alcançar com os indicadores a serem monitorados possibilita o planejamento e execução da coleta de dados ao longo do processo e até mesmo anteriormente, com a possibilidade de identificação do cenário-base, sobre o qual os resultados possam ser comparados. No entanto, muitas vezes isso não é possível, como é o caso da linha Pós-embarque Serviços.

Nesse caso, o esforço de avaliação deve compreender a busca pelo máximo de informações a respeito dos objetivos e metas iniciais do programa, das mudanças e aprimoramentos incorporados ao longo do tempo e suas motivações. Além disso, é preciso lançar mão dos dados disponíveis, em grande parte produzidos por terceiros e que muitas vezes não contemplam todo o período de interesse ou todas as dimensões que se pretende investigar. Também, é preciso tentar contornar a ausência de um cenário-base sobre o qual comparar os efeitos alcançados. Ou seja, o esforço de avaliação deve tentar revelar o quanto a linha contribuiu para avançar nas dimensões historicamente apontadas como relevantes – exportações, empregos e obtenção de divisas –, sem que, contudo, seja possível comparar esse avanço com alguma expectativa previamente definida.

Cadeia de subfornecedores dos projetos: uma vez que a escolha dos subfornecedores ocorre por seleção da empresa exportadora, entre o conjunto de empresas brasileiras disponíveis e sob condições de mercados, os modelos de estimação de impacto sobre essa cadeia devem incluir as hipóteses de seleção sobre observáveis e sobre não observáveis. Nesse caso, espera-se que a incorporação da informação disponível via regressão com controles ou balanceamento por *propensity score* mitigue o viés de seleção com respeito às variáveis observáveis. Contudo, ainda é preciso ter em conta o potencial viés relacionado a características não observáveis. Nesse sentido, a disponibilidade de dados de exportação e emprego para diferentes anos pode contribuir para mitigar tal viés, por meio de técnicas como *diff in diff*.

Adicionalmente, para avaliar se o montante de bens e serviços brasileiros é maior para obras apoiadas, podem-se comparar as exportações de bens e serviços realizadas pelos exportadores beneficiários de Pós-embarque Serviços para países que contaram e que não contaram com financiamento. Também é possível levar em conta, para um mesmo país, os anos em que houve projetos apoiados. Nesses casos, podem ser utilizados dados em painel, com efeitos estimados pelos métodos de regressão com controle para efeito fixo.

Exportação de serviços e balanço de pagamentos: uma vez que a linha Pós-embarque Serviços está disponível para todos os projetos/exportadores que se adequem a seus parâmetros e que são as empresas interessadas que apresentam a solicitação de financiamentos ao BNDES, há um caráter de autosseleção entre as empresas apoiadas e não apoiadas a ser considerado na escolha do método de estimação. Nesse caso, além das metodologias apropriadas para lidar com dados observacionais, propõem-se análises baseadas em métodos quase-experimentais que explorem fatores e eventos exógenos, sejam geográficos, sejam temporais, que afetem a probabilidade de a empresa buscar e/ou obter o apoio.

Por exemplo, a decisão sobre cursar ou não operações de longo prazo de exportação de serviços no Convênio de Pagamentos e Créditos Recíprocos da

Associação Latino-Americana de Integração (CCR/Aladi) é uma decisão que cabe ao país signatário, sendo exógena à empresa exportadora ou ao BNDES. Tal decisão interfere na probabilidade de obtenção de apoio, uma vez que mitiga o risco político e reduz os custos de obtenção do Seguro de Crédito à Exportação (SCE). Logo, seria possível considerar essa diferença regional como uma variável instrumental.

Outro candidato a experimento natural seria a estruturação da linha de crédito a Angola, negociada entre os governos brasileiro e angolano, que estabeleceu previamente garantias que deveriam ser estruturadas por Angola, para obtenção do SCE. É possível testar se houve impacto resultante dessa mudança sobre o desempenho das exportações brasileiras e sobre o balanço de pagamentos bilateral, perante países anteriormente semelhantes e que não contaram com o mesmo instrumento de política comercial.

Por fim, outra possibilidade a ser testada seriam os eventos de interrupção temporária dos desembolsos decorrentes de eventos exógenos, tais como o ocorrido em 2016.

3.3 Dados

Outro aspecto-chave em uma avaliação de impacto/efetividade refere-se à disponibilidade de dados que permitam explorar os objetivos propostos e superar as restrições das metodologias e hipóteses. Tanto o monitoramento de indicadores quanto as avaliações de impacto dependem da disponibilidade de dados relevantes sobre o desempenho de empresas (apoiadas e pertencentes ao grupo de controle) e do setor exportador brasileiro.

O desempenho em relação às exportações de bens e emprego das firmas pode ser obtido por meio de duas fontes tradicionais com cobertura praticamente universal: a Relação Anual de Informações Sociais (Rais) do Ministério do Trabalho e Emprego (MTE), atual Ministério da Economia (ME), e o sistema Alice da Secretaria de Comércio Exterior (Secex). Cabe observar que o acesso aos microdados de ambas as fontes está condicionado à manutenção de acordos de cooperação técnica com os órgãos gestores. No caso do sistema Alice, o BNDES dispõe dos dados com identificação do exportador para o universo das exportações até o ano de 2012. De 2013 em diante, estão disponíveis apenas os dados relativos às empresas com as quais o Banco mantém relacionamento.

Informações econômicas adicionais sobre as empresas (receitas, componentes de custos, nível de investimento etc.) são necessárias para controles das diferenças observáveis entre as empresas, bem como para delinear grupos de controle adequados. Tais informações são coletadas nas pesquisas anuais do Instituto Brasileiro de Geografia e Estatística (IBGE), como a Pesquisa Industrial Anual,

a Pesquisa Anual de Serviços e a Pesquisa Anual da Indústria da Construção. Nesse caso, o uso dos microdados identificados depende do acesso à sala de sigilo do IBGE, cuja concessão depende do envio de projeto e negociação com essa instituição.

Os relatórios de verificação das exportações exigidos no âmbito das operações da linha Pós-embarque Serviços possibilitam maior detalhamento das exportações de bens e serviços associadas aos projetos apoiados. A atualização dessa base é constante e exige harmonização dos formatos em que as informações são reportadas em diferentes operações.

Por fim, dados sobre o mercado internacional de construção podem ser obtidos por meio de publicação anual da *Engineering News-Record* (ENR, 2003; 2004; 2005; 2006; 2007; 2008; 2009; 2010; 2011; 2012; 2013; 2014; 2015; 2016; 2017), que traz o *ranking*, a receita, o setor e o valor dos contratos das maiores empresas de construção do mundo. Informações macroeconômicas sobre os países exportadores e importadores de serviços podem ser obtidas nas bases de dados públicas do Banco Mundial e pela Penn World Table.

Referências

ABADIE, A.; DIAMOND, A.; HAINMUELLER, H. Synthetic control methods for comparative case studies: estimating the effect of California's tobacco control program. *Journal of the American Statistical Association*, [S.l.], v. 105, n. 490, p. 493-505, 2010.

ABADIE, A.; GARDEAZABAL, J. The economic costs of conflict: a case study of the Basque Country. *American Economic Review*, [S.l.], v. 93, n. 1, p. 112-132, 2003.

ENR – ENGINEERING NEWS-RECORD. *The top 225 international contractors*. New York: McGraw Hill, 2003.

ENR – ENGINEERING NEWS-RECORD. *The top 225 international contractors*. New York: McGraw Hill, 2004.

ENR – ENGINEERING NEWS-RECORD. *The top 225 international contractors*. New York: McGraw Hill, 2005.

ENR – ENGINEERING NEWS-RECORD. *The top 225 international contractors*. New York: McGraw Hill, 2006.

ENR – ENGINEERING NEWS-RECORD. *The top 225 international contractors*. New York: McGraw Hill, 2007.

ENR – ENGINEERING NEWS-RECORD. *The top 225 international contractors*. New York: McGraw Hill, 2008.

ENR – ENGINEERING NEWS-RECORD. *The top 225 international contractors*. New York: McGraw Hill, 2009.

ENR – ENGINEERING NEWS-RECORD. *The top 225 international contractors*. New York: McGraw Hill, 2010.

ENR – ENGINEERING NEWS-RECORD. *The top 225 international contractors*. New York: McGraw Hill, 2011.

ENR – ENGINEERING NEWS-RECORD. *The top 225 international contractors*. New York: McGraw Hill, 2012.

ENR – ENGINEERING NEWS-RECORD. *The top 250 international contractors*. New York: McGraw Hill, 2013.

ENR – ENGINEERING NEWS-RECORD. *The top 250 international contractors*. New York: McGraw Hill, 2014.

ENR – ENGINEERING NEWS-RECORD. *The top 250 international contractors*. Troy: BNP Media, 2015.

ENR – ENGINEERING NEWS-RECORD. *The top 250 international contractors*. Troy: BNP Media, 2016.

ENR – ENGINEERING NEWS-RECORD. *The top 250 international contractors*. Troy: BNP Media, 2017.

INTOSAI – THE INTERNATIONAL ORGANISATION OF SUPREME AUDIT INSTITUTIONS. *ISSAI 3000 – Standards and guidelines for performance auditing based on INTOSAI's auditing standards and practical experience*. Stockholm, jul. 2004.

KHANDKER, S. R.; KOOLWAL, G. B.; SAMAD, H. A. *Handbook on impact evaluation: quantitative methods and practices*. Washington D.C.: The World Bank, 2010.

NAJBERG, S.; IKEDA, M. *Modelo de geração de emprego: metodologia e resultados*. Rio de Janeiro: BNDES, 1999. (Textos para Discussão, n. 72).

PINHEIRO, A. C. *et al.* *Incentivos fiscais e creditícios às exportações brasileiras: resultados setoriais para o período 1980-91*. Rio de Janeiro: Ipea, maio 1993. (Texto para Discussão, n. 300).

PINTO, M. A. C. *et al.* *A equalização de taxas de juros como instrumento de promoção das exportações brasileiras de bens de capital no período*

1994-2005. Rio de Janeiro: BNDES, 2008. Disponível em: https://www.bndes.gov.br/SiteBNDES/export/sites/default/bndes_pt/Galerias/Arquivos/conhecimento/especial/equalizacao.pdf. Acesso em: 6 mai. 2010.

TCU – TRIBUNAL DE CONTAS DA UNIÃO. *Manual de auditoria operacional*. Brasília, 2010.

Apêndice metodológico – identificação e estimação do efeito do tratamento¹³

Descrição geral e notação

Um componente básico da notação empregada na literatura de efeito de tratamento é o **modelo de resultados potenciais de Rubin**.¹⁴ Esse modelo postula que, para cada unidade ou observação i , oriunda de uma dada população P , existem dois **resultados potenciais**:

$Y_i(1)$, valor da variável de interesse caso a unidade i esteja sob efeito do tratamento; e

$Y_i(0)$, valor da mesma variável caso i não receba o tratamento.

Além disso, para cada unidade i é possível observar o *status* do tratamento, sinalizado pelo **indicador de tratamento** T_i . De forma análoga aos rótulos dos resultados potenciais, $T_i = 1$ indica que i recebe o tratamento e $T_i = 0$, que não recebe. Por fim, denota-se por X_i um conjunto de **variáveis auxiliares** disponíveis, que se supõe não serem afetadas pelo tratamento.

No modelo de resultados potenciais, o efeito da intervenção sobre a unidade i é dado por $\tau_i = Y_i(1) - Y_i(0)$, desde que: (i) o tratamento esteja bem definido (não há diferentes “versões” do tratamento); e (ii) a atribuição do tratamento a uma unidade não interfira no resultado das demais (por exemplo, não pode haver externalidades ou efeitos de equilíbrio geral substanciais).¹⁵

O **problema fundamental da inferência causal** – termo disseminado por Holland (1986) – consiste no fato de que apenas um dos resultados potenciais é observado para cada indivíduo, mas o cálculo do efeito τ_i do tratamento depende de ambos. Neste ponto da discussão, é útil observar que, mesmo concentrando o foco em agregados, não basta comparar os resultados de unidades tratadas e não tratadas. Por exemplo, a diferença de médias entre os grupos converge probabi-

¹³ Para uma exposição mais ampla e detalhada, sugere-se consultar Khandker, Koolwal e Samad (2010) e Imbens e Wooldridge (2008), que foram utilizados como referência para a elaboração da presente discussão. A aplicação da maioria dos métodos requer o uso de modelos de regressão, que não são cobertos aqui por limite de escopo. Sugere-se consultar Angrist e Pischke (2009), para uma abordagem leve dessa e de outras ferramentas de econometria aplicada, ilustrando alguns dos métodos descritos aqui, e Wooldridge (2010) como referência abrangente.

¹⁴ Assim denominado pela identificação com o estatístico norte-americano Donald Rubin, que o disseminou notoriamente a partir da década de 1970 e contribuiu ativamente para a literatura de inferência estatística causal. As origens do modelo são também atribuídas ao trabalho do matemático e estatístico polonês Jerzy Neyman no início do século XX (o primeiro registro do uso da notação aqui apresentada é datado de 1923). Para uma discussão de origens históricas, ver Imbens e Rubin (2015), cap. 2.

¹⁵ Essas duas condições correspondem à hipótese conhecida como Stable Unit Treatment Value Assumption, cunhada por Rubin em Imbens e Rubin (2015).

listicamente para $E[Y_i(1) | T_i = 1] - E[Y_i(0) | T_i = 0]$. Esse valor pode ser escrito da seguinte forma (somando e subtraindo a média contrafactual $E[Y_i(0) | T_i = 1]$):

$$\{E[Y_i(1) | T_i = 1] - E[Y_i(0) | T_i = 1]\} + \{E[Y_i(0) | T_i = 1] - E[Y_i(0) | T_i = 0]\}$$

em que a primeira parcela consiste no efeito do tratamento sobre as unidades tratadas (uma estatística que frequentemente é o objeto de interesse, como será visto em breve) e a segunda representa a diferença entre os resultados potenciais $Y_i(0)$ das unidades tratadas e não tratadas, uma forma de representar o **viés de seleção** (erro na inferência decorrente das diferenças entre os grupos comparados).

Formalmente, os dados observados são $(Y_i, T_i, X_i)_{i \in S}$, em que $S \subset P$ é uma amostra da população e $Y_i = Y_i(T_i)$ é o **resultado observado** da observação i . Dessa forma, conhece-se apenas $Y_i(1)$ quando $T_i = 1$, e apenas $Y_i(0)$ quando $T_i = 0$. Frequentemente a seguinte representação do resultado observado – cuja verificação é imediata – se mostra útil (para simplificar a notação, será evitado o uso do índice i quando não houver risco de confusão):

$$Y = TY(1) + (1 - T)Y(0) \quad (1)$$

Parâmetro de interesse

A notação apresentada até aqui admite que o efeito do tratamento seja diferente para cada indivíduo. Na prática, deseja-se realizar a inferência sobre alguma quantidade representativa do efeito sobre a população (ou sobre a amostra, quando esta não é representativa da população). Um candidato natural é o valor esperado desse efeito sobre algum subconjunto da população. Duas possibilidades geralmente utilizadas são:

- Efeito médio do tratamento (**ATE**, *average treatment effect*), definido por:

$$\tau_{ATE} = E[\tau_i] = E[Y_i(1) - Y_i(0)]$$

- Efeito médio do tratamento sobre os tratados (**ATT**, *average treatment effect on the treated*), definido por:

$$\tau_{ATT} = E[\tau_i | T_i = 1] = E[Y_i(1) - Y_i(0) | T_i = 1]$$

O ATE representa o efeito esperado da aplicação do tratamento a uma unidade tomada aleatoriamente da população, ou o efeito médio verificado se a intervenção tivesse sido aplicada a toda a população (ou a um subconjunto representativo desta) *vis-à-vis* ao fato de nenhuma unidade ter sido atendida.

O ATT representa o efeito esperado da intervenção sobre uma unidade escolhida aleatoriamente entre as que de fato foram tratadas ou, equivalentemente, o efeito

médio por unidade tratada alcançado pelo programa, dado o mecanismo de seleção atual. Cabe observar que, se a focalização da intervenção foi bem-sucedida, é razoável esperar que o ATT seja maior que o ATE.

Uma generalização do conceito de ATE é o efeito médio local de tratamento (**Late**, *local average treatment effect*), que consiste em uma média ou valor esperado do efeito sobre uma população que não é necessariamente representativa da população de interesse. O ATT é um caso particular de Late.

No contexto de inferência estatística e econometria, o conceito de **identificação** corresponde à capacidade de aprender sobre o parâmetro de interesse pela observação dos dados. Considerando o problema fundamental da inferência causal, é necessário impor hipóteses adicionais sobre o processo gerador dos dados para se obter a identificação. A validade dessas hipóteses, por sua vez, depende da natureza dos dados.

Dados experimentais

Quando a atribuição do tratamento ocorre de forma estritamente aleatória – situação em que se diz que há um **experimento puro** – é lícito supor que os resultados potenciais são estatisticamente independentes da atribuição do tratamento:

$$(Y_i(1), Y_i(0)) \perp T_i \quad (2)$$

Cabe observar que, por construção, nesse caso não há distinção entre o ATT e o ATE (desde que a amostra seja escolhida aleatoriamente a partir da população de interesse).

Outra consequência importante dessa hipótese é que o valor esperado do resultado obtido pelas unidades tratadas é idêntico ao valor esperado do resultado potencial na presença da intervenção:

$$E[Y_i | T = 1] = E[Y_i(1) | T_i = 1] = E[Y_i(1)]$$

em que a primeira igualdade segue da definição de Y_i e a segunda, da hipótese (2). Da mesma forma, o valor esperado do resultado observado para as unidades não tratadas é igual ao valor esperado do valor potencial na ausência da intervenção:

$$E[Y_i | T = 0] = E[Y_i(0) | T_i = 0] = E[Y_i(0)]$$

Em outras palavras, essas unidades constituem um grupo de controle legítimo para fins de cálculo do efeito da intervenção.

Essas relações mostram como, sob a hipótese (2), as quantidades *não observáveis* $E[Y(1)]$ e $E[Y(0)]$ podem ser estimadas via médias dos valores *observados*

de cada grupo. A identificação do ATE nesse contexto segue da observação de que, por linearidade do valor esperado, este pode ser representado pela diferença dessas duas quantidades: $\tau_{ATE} = E[Y(1) - Y(0)] = E[Y(1)] - E[Y(0)]$. De fato, um estimador não viesado do ATE é dado pela simples diferença de médias entre os grupos de controle e tratamento:

$$\hat{\tau}_{ATE} = \frac{\sum_{i|T_i=1} Y_i}{N_t} - \frac{\sum_{i|T_i=0} Y_i}{N_c}$$

em que N_t e N_c são respectivamente o número de unidades no grupo de tratamento e no grupo de controle. No contexto de análise por regressão linear, a mesma inferência pode ser obtida aplicando aos dados o modelo de regressão simples:

$$Y_i = \alpha + \beta T_i + \varepsilon_i$$

em que α e β são parâmetros e ε_i um erro de média zero e não correlacionado com T_i . O estimador de mínimos quadrados ordinários para β será idêntico ao $\hat{\tau}_{ATE}$ descrito anteriormente.

Em alguns casos, a atribuição do tratamento, ainda que realizada de forma experimental, utiliza probabilidades distintas para grupos definidos com base no perfil de variáveis X_i . Nesse caso, diz-se que se trata de um **experimento parcial**, e os métodos descritos a seguir para o contexto mais geral de “seleção em observáveis” se aplicam.

Seleção em observáveis

O contexto de seleção em observáveis é geralmente descrito pelas seguintes hipóteses:

$$(Y_i(1), Y_i(0)) \perp\!\!\!\perp T_i \mid X_i \quad (3)$$

$$\text{Existe } \epsilon > 0 \text{ tal que vale: } \epsilon < P(T_i = 1 \mid X_i = x) < 1 - \epsilon, \text{ para todo } x \quad (4)$$

A hipótese de independência (3), denominada **hipótese de independência condicional**, é menos restritiva que a hipótese (2) – postulada para o caso de experimentos puros –, pois a independência entre os resultados potenciais e a atribuição do tratamento só precisa ocorrer, nesse caso, para unidades semelhantes (com mesmas características observáveis). A independência condicional não é diretamente testável a partir dos dados, isto é, não é possível discernir se ela é válida com base em uma amostra, mesmo depois de um número arbitrariamente grande de observações. Entretanto, é possível discutir se ela é plausível com base em testes indiretos.¹⁶

¹⁶ Ver Imbens e Wooldridge (2008).

Já a hipótese (4), denominada **hipótese de suporte comum** ou de **sobreposição**, essencialmente exige que, para todos os valores das variáveis auxiliares, seja possível observar unidades de controle e tratamento. Essa condição é testável a partir dos dados, e tal verificação é amplamente recomendada pela literatura – ver, por exemplo, Busso, Dinardo e McCrary (2014), Crump *et al.* (2009) ou Imbens (2015). Sua violação reflete na existência de unidades de um dos grupos – tratamento ou controle – com arbitrariamente poucos (ou nenhum) correspondentes no outro e, portanto, na impossibilidade de estimar o respectivo resultado contrafactual. Nesse caso, as alternativas são: (i) por um lado, aceitar hipóteses mais fortes que permitam extrapolação de resultados; e (ii) por outro, abandonar a pretensão a estimar o efeito do tratamento sobre a parte da população em que não há unidades comparáveis.

Há duas formas úteis (no que se refere à exposição) de demonstrar a identificação do ATE sob as hipóteses (3) e (4).

A primeira delas motiva o método de regressão. Observa-se que, desenvolvendo expressões análogas às obtidas com base na hipótese (2), apenas incorporando a natureza condicional da nova hipótese de independência, tem-se:

$$E[Y_i | T = 0, X = x] = E[Y_i(0) | T_i = 0, X = x] = E[Y_i(0)|X = x]$$

$$E[Y_i | T = 1, X = x] = E[Y_i(1) | T_i = 1, X = x] = E[Y_i(1)|X = x]$$

Ou seja, as médias condicionais dos valores observados para as unidades de controle e tratamento equivalem às médias condicionais dos respectivos resultados potenciais. Isso significa que é possível estimar as relações de média condicional dos resultados potenciais:

$$m_1(x) = E[Y_i(1)|X_i = x]$$

$$m_0(x) = E[Y_i(0)|X_i = x]$$

por meio dos dados observados, obtendo estimativas $\hat{m}_1(\cdot)$ e $\hat{m}_0(\cdot)$. Por exemplo, podem-se estimar as regressões lineares:

$$Y(1)_i = \alpha_t + X'_i \beta_t + \varepsilon_{t,i} \quad (5)$$

$$Y(0)_i = \alpha_c + X'_i \beta_c + \varepsilon_{c,i} \quad (6)$$

separadamente para os grupos de tratamento e controle, respectivamente, de modo a obter as estimativas: $\hat{m}_1(x) = \hat{\alpha}_t + \hat{\beta}_t x$ e $\hat{m}_0(x) = \hat{\alpha}_c + \hat{\beta}_c x$.

Além disso, por uma aplicação direta da Lei das Expectativas Iteradas e linearidade à definição do ATE, tem-se:

$$\tau_{ATE} = E[\tau_i] = E[E[\tau_i | X_i = x]] =$$

$$E[E[Y_i(1) - Y_i(0)|X_i = x]] =$$

$$E[E[Y_i(1)|X_i = x]] - E[E[Y_i(0)|X_i = x]]$$

Isso sugere agregar as estimativas $\hat{m}_1(\cdot)$ e $\hat{m}_0(\cdot)$ sobre a amostra inteira, para o ATE:

$$\hat{\tau}_{ATE} = \frac{\sum_{i=1}^N \hat{m}_1(X_i)}{N} - \frac{\sum_{i=1}^N \hat{m}_0(X_i)}{N}$$

ou, seguindo raciocínio análogo para o ATT, sobre o grupo das unidades tratadas:

$$\hat{\tau}_{ATT} = \frac{\sum_{i|T_i=1} \hat{m}_1(X_i)}{N_t} - \frac{\sum_{i|T_i=1} \hat{m}_0(X_i)}{N_t}$$

Considerando estimadores $\hat{m}_1(\cdot)$ e $\hat{m}_0(\cdot)$ lineares em X_i , o estimador para o ATE apresentado anteriormente pode ser também obtido a partir de uma única regressão linear com controles – somando as equações (5) e (6) multiplicadas por T e $1 - T$, respectivamente, e usando a relação (1):

$$Y_i = \alpha_c + (\alpha_t - \alpha_c) T_i + X_i' \beta_c + T_i X_i' (\beta_t - \beta_c) + \varepsilon_{c,i} + T_i (\varepsilon_{t,i} - \varepsilon_{c,i})$$

$$Y_i = \alpha + \delta T_i + X_i' \beta + T_i X_i' \eta + \varepsilon_i \quad (7)$$

sendo possível mostrar que $\hat{\tau}_{ATE} = \hat{\delta} + \bar{X}' \hat{\eta}$, em que \bar{X} é a média das variáveis auxiliares¹⁷ na amostra.¹⁸

Estimadores de **matching** (ou **pareamento**) são motivados da mesma forma que os de regressão. A diferença está na maneira de aproximar $m_1(X_i)$ e $m_0(X_i)$: em vez de estimar as funções $m_1(\cdot)$ e $m_0(\cdot)$ e avaliá-las para diferentes pontos X_i , o método de **matching** simplesmente substitui $m_0(X_i)$ pela média do valor de Y_i para observações do grupo de controle mais próximas a X_i – respectivamente observações tratadas para $m_1(X_i)$. Diz-se que as unidades consideradas similares são *pareadas* (*matched*) pelas variáveis X_i .

Outra forma de identificar o efeito de tratamento usa o conceito de *propensity score*, cuja utilização para fins de inferência causal foi inicialmente proposta por Rosebaum e Rubin (1983). Trata-se de uma função que atribui a probabilidade de receber o tratamento a cada valor das variáveis X_i :

$$p(x) = P[T_i = 1|X_i = x]$$

O *propensity score* pode ser utilizado para obter a média de cada valor potencial para a população, com base em uma média ponderada do valor observado de cada

¹⁷ Alternativamente, a estimação pode ser realizada com as variáveis auxiliares previamente ajustadas pela subtração das respectivas médias amostrais. Isso facilita a análise dos resultados da regressão, uma vez que, nesse caso, $\bar{X} = 0$, de modo que o efeito é evidenciado imediatamente por $\hat{\delta}$.

¹⁸ Em alguns trabalhos se observa o uso da equação simplificada $Y_i = \alpha + \delta T_i + X_i' \beta + \varepsilon_i$, que requer a hipótese adicional de que os efeitos do tratamento não variam entre unidades com características diferentes.

grupo, um método de estimação do efeito de tratamento conhecido como **reponderação**. Intuitivamente, o inverso do *propensity score* em um ponto x representa a proporção (esperada) de unidades com tal perfil na população em comparação ao número de unidades tratadas similares. Ponderado por essa quantidade, o grupo das unidades tratadas torna-se representativo da população.

Formalmente:

$$E\left[\frac{TY}{p(X)}\right] = E\left[E\left[\frac{TY}{p(X)} \middle| X\right]\right] = E\left[\frac{E[TY|X]}{p(X)}\right] = E\left[\frac{E[Y(1)|X] \cdot p(X)}{p(X)}\right] = E[E[Y(1)|X]] = E[Y(1)]$$

em que a primeira e a última igualdade decorrem da Lei das Expectativas Iteradas, e a terceira decorre de:

$$E[TY | X] = E[Y | X, T=1] \cdot P[T=1 | X] + 0 \cdot P[T=0 | X] = E[Y | X, T=1] \cdot p(X)$$

Em suma, a média de valores potenciais $E[Y(1)]$ pode ser expressa como, $E[TY/p(X)]$, que envolve apenas variáveis observáveis. Por um raciocínio análogo, obtém-se $E[(1-T)Y/1-p(X)] = E[Y(0)]$. Substituindo ambas as quantidades no cômputo do ATE:

$$\tau_{ATE} = E\left[\frac{TY}{p(X)}\right] - E\left[\frac{(1-T)Y}{1-p(X)}\right]$$

o que sugere estimar o *propensity score* por alguma função $\hat{p}(\cdot)$ e, em seguida, o efeito médio do tratamento pela diferença de médias ponderadas análoga à expressão anterior:

$$\hat{\tau}_{ATE} = \frac{1}{N} \sum_{i=1}^N \frac{T_i Y_i}{\hat{p}(X_i)} - \frac{1}{N} \sum_{i=1}^N \frac{(1-T_i) Y_i}{1-\hat{p}(X_i)}$$

Os métodos aplicáveis no contexto de seleção em observáveis podem ser implementados de diversas maneiras: para os estimadores de regressão e reponderação, há várias escolhas possíveis para estimadores das funções $m_1(\cdot)$, $m_0(\cdot)$ e $p(\cdot)$, sendo recomendado o uso de métodos flexíveis, como estimadores não paramétricos baseados em *kernel smoothing*, regressão linear local, ou *sieves*. No caso do *matching*, além da escolha do número de pareamentos, é preciso escolher qual o conceito de distância empregado para determinar as unidades a serem pareadas, sendo escolhas populares a métrica de Mahalanobis e o *propensity score*. Dada a multiplicidade de escolhas, a literatura sugere que se realizem testes de robustez, comparando resultados obtidos mediante diferentes especificações.

Além disso, há formas de combinar os três métodos, que em geral oferecem resultados robustos em relação à má especificação de algum deles isoladamente. Destacam-se a estratificação por *propensity score* seguida de regressão – essencialmente aplicar o método de regressão a subamostras definidas por intervalos de *propensity score* –, *matching* com ajuste por regressão – aplicação de regres-

são apenas às unidades pareadas –, e os estimadores “duplamente robustos”, que combinam regressão com reponderação pelo *propensity score* – um exemplo é a estimação da regressão em uma amostra ponderada da forma descrita pelo método da reponderação.¹⁹

Uma recomendação importante no caso de métodos baseados em *propensity score* é analisar o quanto os grupos de controle e tratamento se tornam similares em características observáveis depois da respectiva ponderação por $\frac{1}{p(\cdot)}$ e $\frac{1}{1-p(\cdot)}$. A literatura técnica oferece diversos testes e procedimentos de diagnóstico com essa finalidade. Quando a reponderação falha em equilibrar a distribuição das características entre os grupos, a estimação está sujeita a viés.

Seleção em não observáveis

Quando a atribuição do tratamento depende de fatores não observáveis, a hipótese de independência condicional pode ser violada, e os procedimentos descritos anteriormente deixam de ser válidos. Em outras palavras, mesmo depois de ajustar os desempenhos pelas variáveis X_i , podem restar diferenças nos desempenhos que não dependam apenas do tratamento. A literatura apresenta essencialmente dois tipos de solução para esse problema.

Dados dinâmicos: diferenças em diferenças e controle sintético

Uma das formas de contornar a seleção em fatores não observáveis consiste em explorar informações sobre o desempenho anterior à intervenção. A ideia é que o desempenho prévio reflete os fatores não observáveis, possibilitando então descontar esse efeito das diferenças verificadas depois do tratamento, ou comparar unidades com não observáveis similares detectados justamente pelo desempenho histórico similar.

Diferenças em diferenças

Quando as variáveis que determinam a seleção não são observáveis, a hipótese de independência condicional $(3) - (Y_i(1), Y_i(0)) \perp\!\!\!\perp T_i \mid X_i$ – não é mais válida e, por conseguinte, estimadores como os de regressão e *matching* deixam de fornecer resultados não viesados para o impacto. No entanto, sempre que os grupos de controle e tratamento tiverem trajetórias paralelas no tempo, na ausência do tratamento (diferenças invariantes no tempo), o método de diferenças em diferenças (*diff in diff*) pode ser utilizado.

¹⁹ Para uma apresentação dos dois primeiros métodos – inicialmente propostos por Rubin e colaboradores –, ver Imbens e Rubin (2015). A literatura de estimação duplamente robusta foi originada pelo epidemiologista James Robins e colaboradores; ver, por exemplo, Robins e Rotnitzky (1995), Scharfstein, Rotnitzky e Robins (1999) e – para um tratamento recente – Athey *et al.* (2017).

Esse método tem como vantagem o fato de não exigir que os grupos de controle e tratamento tenham o mesmo resultado na ausência da intervenção (sejam diretamente comparáveis) ou que as variáveis não observadas permaneçam constantes no tempo, mas somente que tenham a mesma tendência. Em outras palavras, o viés de seleção ou a diferença entre os dois grupos deve permanecer invariante no tempo, na ausência do tratamento. A ação do tratamento é, então, a de modificar essa tendência. Logo, para que seu impacto seja calculado, basta comparar as informações dos dois grupos antes e depois da intervenção. Para tal, não são necessários dados em painéis. Dados transversais para, ao menos, dois períodos (antes e depois da intervenção) podem ser suficientes.

Relembrando a especificação paramétrica (7), válida para o caso de seleção em observáveis, e ignorando a heterogeneidade do efeito do tratamento de acordo com as características (ou seja, supondo $\eta = 0$):

$$Y_i = \alpha + \delta T_i + X_i' \beta + \varepsilon_i$$

a estimação dessa equação por regressão linear produz um valor consistente para $\hat{\tau}_{ATE}$ (no caso, igual ao estimador do parâmetro δ) quando a hipótese (3) é válida, pois esta implica que as variáveis T_i e ε_i são não correlacionadas, dado o valor de X_i . No entanto, se há seleção em não observáveis, a hipótese (3) não se verifica e o estimador pode apresentar viés de sentido e magnitude arbitrários.

A estratégia de identificação em tela supõe que a fonte de correlação entre T_i e ε_i consiste em algum fator específico às unidades que evolui paralelamente entre os grupos ao longo do tempo. Nesse contexto, os dados são descritos por $(Y_{ij}, T_i)_{i \in S, j \in \{0, 1, 2, \dots, T\}}$, e o índice j representa o período de tempo. É conveniente, para fins de exposição, considerar que o tratamento é aplicado em um único período, de forma irreversível, de modo a afetar os resultados a partir desse período. Pela hipótese descrita, o erro ε_{ij} se decompõe da seguinte forma:

$$\varepsilon_{ij} = \phi_i + \theta_j + u_{ij}$$

em que $E[u_{ij}] = 0$ e $u_{ij} \perp T_i$ e, portanto, $E[u_{ij} | T_i = 1] = E[u_{ij} | T_i = 0] = 0$.

Seguindo a descrição de Cameron e Trivedi (2005), dada uma intervenção no período t , a variação do resultado da unidade i pode ser escrita como: $[Y_{it} - Y_{i(t-1)} | T_i = 1]$, se ela pertencer ao grupo de tratamento, ou $[Y_{it} - Y_{i(t-1)} | T_i = 0]$, se ela pertencer ao grupo de controle. Logo, a diferença entre as duas evoluções é dada por $[Y_{it} - Y_{i(t-1)} | T_i = 1] - [Y_{it} - Y_{i(t-1)} | T_i = 0]$, que é a **medida de diferenças em diferenças**.

Sob a hipótese anterior, o efeito do tratamento é dado pela agregação dessa medida:

$$E[Y_{it} - Y_{i(t-1)} | T_i = 1] - E[Y_{it} - Y_{i(t-1)} | T_i = 0]$$

$$\begin{aligned}
&= \{E[Y_{it} | T_i = 1] - E[Y_{it} | T_i = 0]\} - \{E[Y_{i(t-1)} | T_i = 1] - E[Y_{i(t-1)} | T_i = 0]\} \\
&= \{\alpha + \delta + \theta_t + E[X_i' \beta + \phi_i | T_i = 1] - (\alpha + \theta_t + E[X_i' \beta + \phi_i | T_i = 0])\} - \\
&\quad \{\alpha + \theta_{t-1} + E[X_i' \beta + \phi_i | T_i = 1] - (\alpha + \theta_{t-1} + E[X_i' \beta + \phi_i | T_i = 0])\} \\
&= \{\delta + E[X_i' \beta + \phi_i | T_i = 1] - E[X_i' \beta + \phi_i | T_i = 0]\} - \\
&\quad \{E[X_i' \beta + \phi_i | T_i = 1] - E[X_i' \beta + \phi_i | T_i = 0]\} \\
&= \delta = \tau_{ATE}
\end{aligned}$$

A expressão da segunda linha deixa claro que bastam amostras repetidas dos grupos de controle e tratamento (sendo desnecessário que estes sejam compostos pelas mesmas unidades nos dois períodos). O estimador, no caso, consiste no análogo amostral:

$$\hat{\tau}_{did} = \left\{ \frac{\sum_{T_i=1, j=t} Y_{ij}}{N_{1,t}} - \frac{\sum_{T_i=0, j=t} Y_{ij}}{N_{0,t}} \right\} - \left\{ \frac{\sum_{T_i=1, j=t-1} Y_{ij}}{N_{1,t-1}} - \frac{\sum_{T_i=0, j=t-1} Y_{ij}}{N_{0,t-1}} \right\}$$

Na linguagem de regressão, a especificação correspondente é a seguinte:

$$Y_{ij} = \alpha + \phi T_i + \theta pós_j + \delta(T_i \times pós_j) + u_{ij}$$

em que $pós_j$ é a variável binária que indica que a observação ocorre depois do tratamento, ou seja: $pós_j = 0$, se $j = t - 1$; e $pós_j = 1$, se $j = t$. É importante notar que, nesse caso, o coeficiente do termo relacionado ao tratamento, ϕ , apenas captura a diferença entre os grupos pré-tratamento. O termo associado ao período, por sua vez, reflete a tendência temporal ($\theta_t - \theta_{t-1}$), na descrição do modelo desta seção). O estimador do efeito de tratamento é dado pelo coeficiente do termo de interação: $\hat{\tau}_{did} = \hat{\delta}$.

Se as mesmas unidades são observadas nos dois períodos, o ATE pode também ser estimado por regressão, bastando, nesse caso, controlar diretamente para os resultados pré-tratamento das unidades. A generalização para vários períodos dessa abordagem consiste nas técnicas para **dados em painel** (nome genérico para a estrutura de dados em que é possível observar as mesmas unidades em diferentes pontos no tempo), que permitem isolar os efeitos fixos individuais com mais precisão. O estudo de dados em painel requer conhecimentos avançados de modelos de regressão e técnicas de inferência estatística; uma discussão ampla dessas técnicas está disponível em Wooldridge (2010).

Controle sintético

Uma alternativa que explora de forma mais complexa a dinâmica prévia ao tratamento é a metodologia de **controle sintético**. Essa metodologia, apresentada em Abadie

e Gardeazabal (2003) e detalhada em Abadie, Diamond e Hainmueller (2010),²⁰ é adequada para análise de efeitos em nível agregado e que afetam um número relativamente pequeno de unidades. Consiste na construção de uma unidade de controle sintética com base na média ponderada das principais características das unidades de controle disponíveis, de forma que tal unidade sintética aproxime as características relevantes da unidade tratada, anteriormente à intervenção analisada. Assim, o impacto da intervenção sobre a unidade tratada pode ser verificado pela diferença entre esta e a unidade de controle sintético, no período posterior à intervenção.

Considera-se um conjunto de $I + 1$ unidades, das quais uma recebe o tratamento ($i = 1$, por conveniência) e as demais são potenciais unidades de controle ($i = 2, \dots, I + 1$).²¹ As unidades são observadas nos períodos $t = 1, \dots, T$, com a intervenção ocorrendo em (ou afetando os resultados a partir de) $T_0 + 1$, de forma que $1, \dots, T_0$ sejam os períodos pré-intervenção e $T_0 + 1, \dots, T$ sejam os períodos pós-intervenção. Além da variável de interesse, observa-se, ainda, para cada unidade, U_i , um vetor ($r \times 1$) de variáveis auxiliares, as quais consistem em um conjunto de preditores para a variável de resultado a ser avaliada.

Propõe-se estimar o efeito sobre a unidade tratada $\tau_{1t} = Y_{1t}(1) - Y_{1t}(0)$, para os períodos $T_0 + 1, \dots, T$. Como o resultado potencial da unidade $i = 1$ na ausência do tratamento é contrafactual, recorre-se a uma combinação dos desempenhos das demais unidades. Ao fixar um vetor de pesos $W = (w_2, w_3, \dots, w_{I+1})$, com w_i não negativos e $\sum_{i=2}^{I+1} w_i = 1$, para as potenciais unidades de controle, define-se um controle sintético arbitrário para a unidade tratada. Assim, o método de controle sintético propõe que os resultados contrafactuais $Y_{1t}(0)$, para $t > T_0$, sejam estimados por alguma combinação $\hat{Y}_{1t}(0) = \sum_{i=2}^{I+1} w_i^* Y_{it}$, com pesos W^* .

O ponto crucial da técnica é a escolha dos pesos. Propõe-se que o controle sintético seja tão similar quanto possível à unidade tratada em relação a um conjunto de valores X_i , composto pelas características U_i e combinações dos resultados pré-tratamento, definidas por M vetores $\bar{Y}_i^{K_m} = \sum_{s=1}^{T_0} k_{m,s} Y_{is}$, para $m = 1, 2, \dots, M$. É essa combinação linear dos resultados pré-intervenção que será utilizada para controlar os fatores comuns não observáveis que variam ao longo do tempo. Formalmente, é demonstrado que, se o número de períodos pré-tratamento incluídos na análise for suficientemente grande, o erro se torna arbitrariamente pequeno.

Em suma, sejam $X_i = (U_i, \bar{Y}_i^{K_1}, \dots, \bar{Y}_i^{K_M})'$, para $i = 1, 2, \dots, I$, vetores ($k \times 1$), $k = r + M$, contendo as características observadas e combinações de resultados pré-tratamento de cada unidade, e $X_0 = (X'_2, X'_3, \dots, X'_I)$, a matriz ($k \times I$) contendo as características de todos os controles potenciais, a implementação do controle

²⁰ Os mesmos autores produziram um pacote de comandos estatísticos computacionais a ser utilizado no programa R (*Synth package*), o qual detalham em Abadie, Diamond e Hainmueller (2011). O citado artigo traz uma apresentação mais sucinta do modelo, a qual foi utilizada como base para a descrição desta seção.

²¹ Os autores chamam esse subconjunto de “*donor pool*”.

sintético consiste em escolher o vetor de pesos W^* que minimize alguma distância $\|X_1 - X_0 W\|$ entre as características da unidade tratada e do controle sintético. Geralmente, usa-se a distância:

$$\|X_1 - X_0 W\|_V = \sqrt{(X_1 - X_0 W)' V (X_1 - X_0 W)}$$

em que V é uma matriz $(k \times k)$ simétrica positiva semidefinida. A escolha de V determina os pesos dados às diferenças em cada dimensão e afeta a precisão (erro quadrático médio), mas não a consistência do estimador. A escolha ótima (que confere menor erro quadrático médio) depende de valores desconhecidos, mas que podem ser estimados com base nos dados.

Contextos quase-experimentais: variáveis instrumentais e *regression discontinuity design*

Variáveis instrumentais

Outra forma de evitar viés por seleção em não observáveis é encontrar fatores de atribuição aleatória de tratamento, gerando contextos comparáveis aos dos experimentos. Tipicamente, essa construção é obtida pelo método de **variáveis instrumentais**.

No contexto de variáveis instrumentais, está disponível uma variável, denominada **instrumento** para o indicador de tratamento, denotada por Z , que afeta a seleção para o tratamento, mas não os resultados potenciais. Supondo Z binária, valendo 1 caso haja indução ao tratamento e 0, caso contrário, e aplicando notação análoga à do modelo de resultados potenciais para a relação entre Z_i e T_i , postula-se a existência de condições de participação potenciais $T_i(0)$ e $T_i(1)$ de acordo com o valor assumido pelo instrumento. Os dados observados consistem em $(Y_i, T_i, Z_i)_{i \in S}$, em que $T_i = T_i(Z_i)$.

Existem quatro comportamentos, ou padrões de resposta, logicamente possíveis:

$$C_i = \begin{cases} n \text{ (never-taker)}, & \text{se } T_i(0) = T_i(1) = 0 \\ c \text{ (complier)}, & \text{se } T_i(0) = 0 \text{ e } T_i(1) = 1 \\ d \text{ (defier)}, & \text{se } T_i(0) = 1 \text{ e } T_i(1) = 0 \\ a \text{ (always-taker)}, & \text{se } T_i(0) = T_i(1) = 1 \end{cases}$$

Por exemplo: diz-se que uma unidade é *complier* (“obediente”) se ela participa do tratamento quando induzida (isto é, $T_i(1) = 1$) e não participa quando não induzida ($T_i(0) = 0$). Cabe observar que, a princípio, não é possível deduzir o tipo de cada unidade com base nos dados observados: se, por exemplo, $T_i = 1$ e $Z_i = 1$, é possível que i seja *complier* ou *always-taker*.

Para fins de identificação, é necessário impor a restrição de **monotonicidade**, ou ausência de *defiers*:

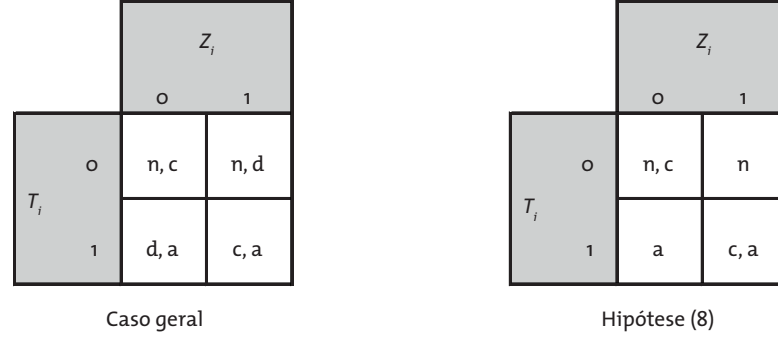
$$C_i \in \{n, c, a\} \tag{8}$$

e a **hipótese de independência**:

$$(Y_i(1), Y_i(0), T_i(1), T_i(0)) \perp Z_i \quad (9)$$

A Figura A1 ilustra os comportamentos possíveis e as restrições implicadas pela hipótese (8).

Figura A1. Hipótese de monotonicidade



Fonte: Elaboração própria.

A hipótese (9), por sua vez, assegura que o instrumento é independente do comportamento e dos resultados potenciais de cada indivíduo. Quando as duas hipóteses são válidas, os conjuntos de observações definidos por $S^1 = \{i \in S | Z_i = 1\}$ e $S^0 = \{i \in S | Z_i = 0\}$ diferem apenas na medida em que os *compliers* estão sob tratamento no primeiro (logo assumem $T = 1$ e $Y = Y(1)$), mas não no segundo (logo assumem $T = 0$ e $Y = Y(0)$).

Assim, em ambos os conjuntos, a proporção esperada de *compliers* é idêntica à observada na população em geral: $\pi_c = P[C_i = c] = E[T_i | Z_i = 1] - E[T_i | Z_i = 0]$ (em razão da hipótese (9), por independência entre Z_i e C_i). Além disso, a diferença do desempenho médio dos grupos é dada apenas pela variação da condição do tratamento dos *compliers*:

$$E[Y_i | Z_i = 1] - E[Y_i | Z_i = 0] = \pi_c \cdot E[Y_i(1) - Y_i(0) | C_i = c]$$

Portanto, é possível identificar o efeito de tratamento local para os *compliers* como:

$$\tau_{IV} = E[Y_i(1) - Y_i(0) | C_i = c] = \frac{E[Y_i | Z_i = 1] - E[Y_i | Z_i = 0]}{E[T_i | Z_i = 1] - E[T_i | Z_i = 0]}$$

e estimá-lo por meio do análogo amostral:

$$\hat{\tau}_{IV} = \frac{\frac{\sum_{i \in S^1} Y_i}{N^1} - \frac{\sum_{i \in S^0} Y_i}{N^0}}{\frac{\sum_{i \in S^1} T_i}{N^1} - \frac{\sum_{i \in S^0} T_i}{N^0}}$$

em que N^1 e N^0 são o número de observações em S^1 e S^0 .

Considerando-se a ideia básica de variável instrumental, estão disponíveis diversas extensões, como instrumento com vários níveis (em vez de binário) e modelos com heterogeneidade por fatores observáveis (variáveis X_i).

É importante notar que, para obter o ATE ou o ATT no contexto de variáveis instrumentais, são necessárias hipóteses adicionais que permitam a extrapolação do efeito identificado para os *compliers*. Nesse sentido, a literatura sugere testes indiretos que podem revelar se essa extrapolação é plausível, como verificar em que medida os *always-takers* se assemelham aos *compliers* que recebem o tratamento (analogamente para *never-takers* e *compliers* não tratados).

Regression discontinuity design

Uma estratégia relacionada à de variáveis instrumentais é a exploração de um ***regression discontinuity design (RDD)***, que consiste em um desenho quase-experimental em que a atribuição para o tratamento varia descontinuamente conforme o valor de uma determinada característica. Um exemplo típico é a existência de um limite de elegibilidade imposto pela política analisada. A variável para a qual se verifica a descontinuidade chama-se *running variable*, *forcing variable*, ou *assignment variable* e será denotada por Z ; o limiar em que a descontinuidade ocorre será representado por Z^* .

O caso mais simples é o denominado *sharp design*, em que a descontinuidade é determinante do tratamento, ou seja:

$$P[T_i = 1|Z_i] = \begin{cases} 0, & \text{se } Z_i < z^* \\ 1, & \text{se } Z_i > z^* \end{cases}$$

Quando a relação entre a descontinuidade e a atribuição de tratamento é probabilística, tem-se:

$$p^- = \lim_{z \uparrow z^*} P[T_i = 1|Z_i = z] \neq \lim_{z \downarrow z^*} P[T_i = 1|Z_i = z] = p^+$$

que torna o contexto análogo ao caso geral de variáveis instrumentais com proporção $p^+ - p^-$ de *compliers*. Para um tratamento unificado dos casos, basta observar que o *sharp design* é o caso-limite em que $p^+ = 1$ e $p^- = 0$.

A hipótese de identificação consiste na continuidade das funções de média condicional dos valores potenciais em relação à variável Z : $m_1(z) = E[Y_i(1) | Z_i = z]$ e $m_0(z) = E[Y_i(0) | Z_i = z]$. Nessas condições:

$$E[Y_i(1)|Z_i = z^*] = \lim_{z \uparrow z^*} m_1(z) = \lim_{z \downarrow z^*} m_1(z)$$

$$E[Y_i(0)|Z_i = z^*] = \lim_{z \uparrow z^*} m_0(z) = \lim_{z \downarrow z^*} m_0(z)$$

Consequentemente, nesse contexto, de forma similar ao que acontece para variáveis instrumentais, a diferença entre os desempenhos médios verificados em vizinhanças arbitrariamente pequenas de valores de Z_i menores e maiores que z^* será devida à variação do tratamento. A quantidade a ser estimada é:

$$\tau_{RDD} = \frac{\lim_{z \downarrow z^*} E[Y_i | Z_i = z] - \lim_{z \uparrow z^*} E[Y_i | Z_i = z]}{\lim_{z \downarrow z^*} E[T_i | Z_i = z] - \lim_{z \uparrow z^*} E[T_i | Z_i = z]}$$

O estimador consiste em:

$$\hat{\tau}_{RDD} = \frac{\hat{m}^+(z^*) - \hat{m}^-(z^*)}{\hat{p}^+(z^*) - \hat{p}^-(z^*)}$$

em que: $\hat{m}^+(\cdot)$ e $\hat{m}^-(\cdot)$ são estimativas para a média condicional do resultado observado com respeito a Z , considerando, no primeiro caso, somente pontos à direita do limiar ($Z < z^*$), e no segundo, à esquerda; e $\hat{p}^+(\cdot)$ e $\hat{p}^-(\cdot)$ são estimativas para a probabilidade de tratamento condicional a Z , considerando, respectivamente, pontos à direita e à esquerda do limiar. Os métodos para a estimação dessas funções devem priorizar a precisão no ponto de fronteira $Z = z^*$. Assim, as técnicas de suavização por *kernel* e regressão linear local, que têm como base ponderações que magnificam o peso de observações próximas a cada ponto estimado, são mais indicadas que o método de *sieves*, que tenta ajustar globalmente uma tendência aos dados.

No contexto de RDD, a posição relativa à descontinuidade funciona como instrumento para T . Assim como no caso de variáveis instrumentais, o RDD permite apenas identificar um efeito de tratamento local, a saber, aquele válido para unidades sensibilizadas pelo instrumento, ou seja, aquelas com o valor da *forcing variable* próximo ao limiar z^* e que sejam afetadas pelo limiar (isto é, os *compliers*).²²

A hipótese de continuidade dos valores potenciais não pode ser diretamente testada, mas recomenda-se verificar se outras variáveis não têm comportamento descontínuo. Caso isso ocorra, é provável que outros mecanismos além da relação causal com o tratamento possam afetar os resultados observados, o que seria uma evidência desfavorável à validade da metodologia (IMBENS; WOOLDRIDGE, 2008). Outra recomendação prática é avaliar se a distribuição da *running variable* é contínua perto do limiar. Caso isso não se verifique, é possível que haja manipulação dessa variável, o que também invalida os resultados obtidos via RDD.

²² Formalmente, a inferência é válida de forma exata apenas para o (estrito) grupo de unidades com $Z_i = z^*$ e extensível de forma aproximada para valores próximos, por continuidade.

Referências

- ABADIE, A.; DIAMOND, A.; HAINMUELLER, H. Synth: an R Package for synthetic control methods in comparative case studies. *Journal of Statistical Software*, [S.l.], v. 42, n. 13, p. 1-17, 2011.
- ABADIE, A.; DIAMOND, A.; HAINMUELLER, H. Synthetic control methods for comparative case studies: estimating the effect of California's tobacco control program. *Journal of the American Statistical Association*, [S.l.], v. 105, n. 490, p. 493-505, 2010.
- ABADIE, A.; GARDEAZABAL, J. The economic costs of conflict: a case study of the Basque Country. *American Economic Review*, Pittsburgh, v. 93, n. 1, p. 112-132, 2003.
- ANGRIST, J. D.; PISCHKE, J. *Mostly harmless econometrics: an empiricist's companion*. Princeton: Princeton University Press, 2009.
- ATHEY, S. *et al.* Estimating average treatment effects: Supplementary analyses and remaining challenges. *American Economic Review*, Nashville, v. 107, n. 5, p. 278-281, 2017.
- BUSO, M.; DINARDO, J.; MCCRARY, J. New Evidence on the finite sample properties of propensity score reweighting and matching estimators. *Review of Economics and Statistics*, Cambridge, v. 96, n. 5, p. 885-897, 2014.
- CAMERON, A. C.; TRIVEDI, P. K. *Microeconometrics: methods and applications*. New York: Cambridge University Press, 2005.
- CRUMP, R. K. *et al.* O. A. Dealing with limited overlap in estimation of average treatment effects. *Biometrika*, Oxford, v. 96, n. 1, 2009.
- HOLLAND, P. W. Statistics and Causal Inference. *Journal of the American Statistical Association*, [S.l.], v. 81, n. 396, p. 945-960, 1986.
- IMBENS, G. W. Matching methods in practice: tree examples. *The Journal of Human Resources*, Madison, v. 50, n. 2, p. 373-419, 2015.
- IMBENS, G. W. ; RUBIN, D. B. *Causal inference for statistics, social, and biomedical sciences: an introduction*. New York: Cambridge University Press, 2015.
- IMBENS, G. W.; WOOLDRIDGE, J. M. Recent developments in the econometrics of program evaluation. *Journal of Economic Literature*, Nashville, v. 47, n. 1, p. 5-86, 2008.

KHANDKER, S. R.; KOOLWAL, G. B.; SAMAD, H. A. *Handbook on impact evaluation: quantitative methods and practices*. Washington D.C: The World Bank, 2010.

ROBINS, J. M.; ROTNITZKY, A. Semiparametric efficiency in multivariate regression models with missing data. *Journal of the American Statistical Association*, v. 90, n. 1, p. 122-129. New York: American Statistical Association, 1995.

ROSENBAUM, P. R.; RUBIN, D. B. The central role of the propensity score in observational studies for causal effects. *Biometrika*, Oxford, v. 70, n. 1, p. 41-55, 1983.

SCHARFSTEIN, D. O.; ROTNITZKY, A.; ROBINS, J. M. Adjusting for nonignorable drop-out using semiparametric nonresponse models. *Journal of the American Statistical Association*, New York, v. 94, n. 448, p. 1.096-1.120, 1999.

WOOLDRIDGE, J. M. *Econometric analysis of cross section and panel data*. 2. ed. Cambridge, Mass: MIT Press, 2010.

Coordenação Editorial

Gerência de Editoração e Memória
do BNDES

Projeto Gráfico

Fernanda Costa e Silva

Produção Editorial

Expressão Editorial

Editoração Eletrônica

Expressão Editorial

Editado pelo
Departamento de Comunicação
Junho de 2019



MINISTÉRIO DA
ECONOMIA



PÁTRIA AMADA
BRASIL
GOVERNO FEDERAL

www.bndes.gov.br